Small Business Innovation Research/Small Business Tech Transfer

Enabling High Efficiency I/O Performance on High-End Computing Systems and Data Stores with HDF5, Phase I

Completed Technology Project (2018 - 2019)

Project Introduction

Akadio proposes to develop *H5 Hermes*, an enhancement to HDF5 to enable high efficiency I/O performance for HDF5 applications that perform many related reads and writes across large data structures. Such I/O operations are often a serious bottleneck for large scale simulations, resulting in I/O bottlenecks that significantly degrade overall performance.

There are existing HDF5-based solutions similar to what we propose here. However, these solutions are either one-off implementations or implemented for specific computing environments, or both, and thus do not have the broad applicability that a general HDF5 solution would have. With its proven record for scalability, HDF5 is the data store of choice for many of the most important applications at NASA, the DOE, and other high-end computing (HEC) facilities. Thus, an HDF5 based solution would become available immediately, with minimal modification, to many existing applications, and would be extremely easy to adapt in support of future applications.

H5 Hermes directly addresses the objectives of the solicitation. It will accelerate the integration of current and future high-end computing systems and data stores by enabling I/O processing to keep up with the demands of increasingly high-resolution simulations on massively parallel systems. Thus, H5 Hermes helps achieve four of the five objectives of solicitation S5.01, namely to:

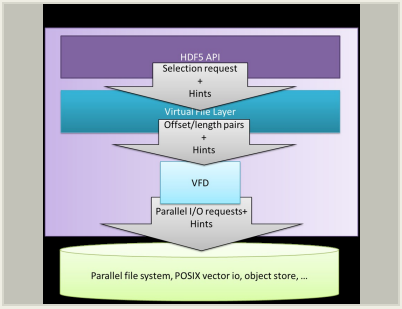
- Minimize the supercomputer user's total time-to-solution.
- Increase the achievable scale and complexity of computational analysis, data ingest, and data communications.
- Reduce the cost of providing a given level of supercomputing performance for NASA applications.
- Enhance the efficiency and effectiveness of NASA's supercomputing operations and services.

H5 Hermes will also improve the ability to efficiently retrieve many pieces of related data from very large cloud-based data stores. This is another NASA priority, though not part of the S5.01 solicitation.

Anticipated Benefits

A wide range of NASA applications use HDF5 or one based on HDF5, such as netCDF4 and CGNS. In earth science, the GMAO and NCCS use HDF5 or netCDF4 for HEC simulations. In flight-dynamics simulations, NASA uses HDF5, often on high end systems requiring high speed parallel I/O. Many astrophysics and astronomy codes, such as ENZO and FLASH use HDF5 for simulation, modeling and analysis, on HEC systems. The Astrophysics Source Code Library lists several codes that use HDF5 and run on HEC systems.

Climate and weather modeling codes use HDF5 or netCDF4 . E.g. Ocean-Land-Atmosphere Model. Astrophysics codes (e.g. ENZO and FLASH) use HDF5 for



Enabling High Efficiency I/O Performance on High-End Computing Systems and Data Stores with HDF5, Phase I

Table of Contents

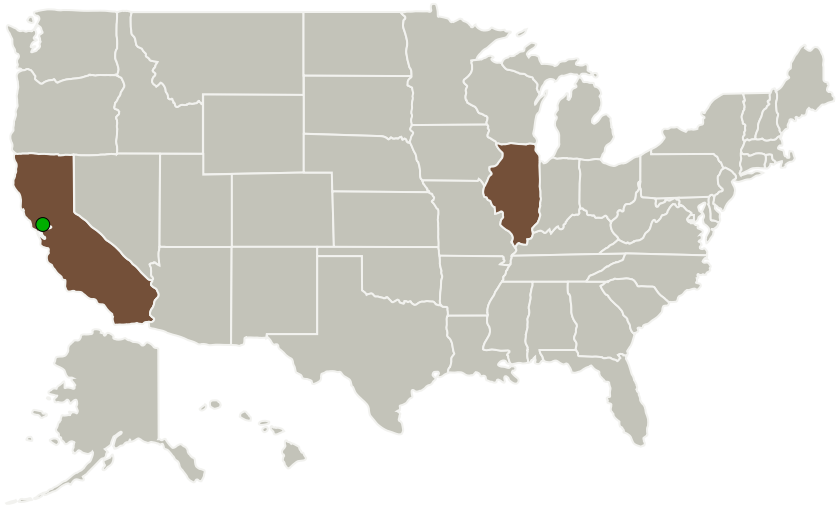
Project Introduction	1
Anticipated Benefits	1
Primary U.S. Work Locations and Key Partners	2
Project Transitions	2
Organizational Responsibility	2
Project Management	2
Technology Maturity (TRL)	2
Images	3
Technology Areas	3
Target Destinations	3

Enabling High Efficiency I/O Performance on High-End Computing Systems and Data Stores with HDF5, Phase I

Completed Technology Project (2018 - 2019)

simulation, and push I/O limits on HEC systems. All major aerospace companies, the finance industry, oil and gas industry, and many others, use HDF5 in simulations and other analysis on HEC systems. HDF5 is the most-used format for applications in the DOE laboratories, largely because is performs so sell on the world's biggest and fastest computers.

Primary U.S. Work Locations and Key Partners



Organizations Performing Work	Role	Type	Location
Akadio, Inc.	Lead Organization	Industry	Champaign, Illinois
Ames Research Center(ARC)	Supporting Organization	NASA Center	Moffett Field, California

Primary U.S. Work Locations	
California	Illinois

Project Transitions

July 2018: Project Start

Organizational Responsibility

Responsible Mission Directorate:

Space Technology Mission Directorate (STMD)

Lead Organization:

Akadio, Inc.

Responsible Program:

Small Business Innovation Research/Small Business Tech Transfer

Project Management

Program Director:

Jason L Kessler

Program Manager:

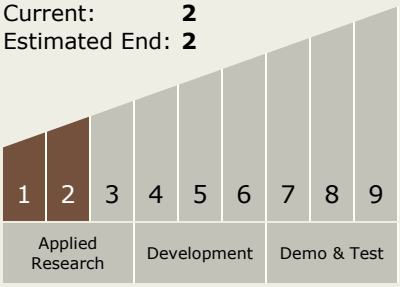
Carlos Torrez

Principal Investigator:

Michael Folk

Technology Maturity (TRL)

Start: **1**
Current: **2**
Estimated End: **2**



Enabling High Efficiency I/O Performance on High-End Computing Systems and Data Stores with HDF5, Phase I

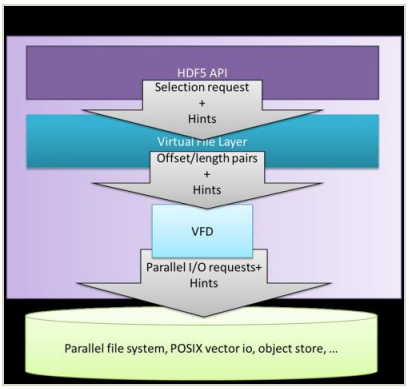
Completed Technology Project (2018 - 2019)

February 2019: Closed out

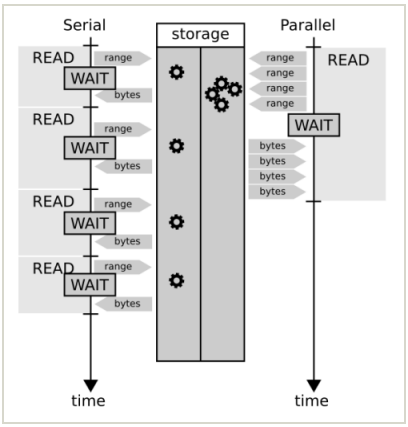
Closeout Documentation:

- Final Summary Chart*(https://techport.nasa.gov/file/141039)*

Images



Briefing Chart Image
Enabling High Efficiency I/O Performance on High-End Computing Systems and Data Stores with HDF5, Phase I *(https://techport.nasa.gov/image/128823)*



Final Summary Chart Image
Enabling High Efficiency I/O Performance on High-End Computing Systems and Data Stores with HDF5, Phase I *(https://techport.nasa.gov/image/132654)*

Technology Areas

Primary:

- TX11 Software, Modeling, Simulation, and Information Processing
  - TX11.4 Information Processing
    - TX11.4.4 Collaborative Science and Engineering

Target Destinations

Earth, Others Inside the Solar System, Outside the Solar System